

A Minority Report: Fundamental Concerns about the CDC Meta-analysis of Group-based Interventions to Prevent Adolescent Pregnancy, HIV, and Other STIs

by Irene Ericksen, M.S. & Danielle Ruedt, M.P.H.,
Members of the CDC Community Guide
Adolescent Sex Behavior Coordination Team External Partners consultant panel

November 6, 2009

I. Summary Statement

We are grateful to have been involved in the Centers for Disease Control and Prevention's research study, "Group-based Interventions to Prevent Adolescent Pregnancy, HIV, and Other Sexually Transmitted Infections: A Systematic Review and Meta-analysis." This was an important effort to summarize the outcome research on sex education in America. We acknowledge the careful work by the Community Guide research staff and their willingness to consider all viewpoints when making decisions about the research. ***What follows represents our minority opinion as members of this study's External Partners consultant panel and does not represent the views of the CDC or the Adolescent Sex Behavior Coordination Team.***

There are serious limitations to this meta-analysis study that cause us to take exception to the Recommendation Statements issued by the CDC's Task Force on Community Preventive Services about the study results. These limitations lead us to conclude that the statements about the general effectiveness of the comprehensive risk reduction (CRR) strategy are not warranted by the data. They overstate the likelihood that any single CRR program will be effective at protecting the sexual health of adolescents, especially the school-based programs, which are the focus of the public policy debate about sex education and impact the future health of millions of adolescents across the country. The CDC recommendations also fail to acknowledge the effectiveness of abstinence education (AE) programs at reducing teen sexual activity, and make comparative effectiveness claims about CRR versus AE that are based on weakly supported assumptions.

The Task Force has made public its Recommendation Statements without also making available to the public the full set of study findings upon which the recommendations are based—both supporting and otherwise. The reason given for this decision is that the data from the study has not yet been scientifically cleared by the CDC for release to the public. However, this policy prevents the public from scrutinizing the body of evidence underlying the CDC Task Force Recommendations in the same time frame in which the CDC recommendations will influence the decisions of policymakers and public health professionals. (Having the opportunity to examine this evidence is particularly important in the current climate of controversy and politicization that surrounds the public policy debate about sex education in America.)

Additionally, this policy prevents the inclusion in this minority report of most of the scientific data that supports and illustrates the concerns described herein. While the meta-analysis data has been made available to us as members of the Adolescent Sex Behavior Coordination Team, at

this point the Recommendation Statements can only be addressed in a qualitative manner here, without reporting any quantitative data that has not been released to the public.

Given the public's inability to fully examine the scientific evidence for the Task Force Recommendations, this minority report offers evidence that there are alternative interpretations of the meta-analysis results which have scientific merit. These observations and assertions are offered in good faith, with the intent of promoting an open dialogue about the best interpretation of the results of this meta-analysis study.

II. Fundamental Concerns and Supporting Rationale

1. The conclusion of general CRR effectiveness is not supported by the totality of data.

The Recommendation Statement asserts that the comprehensive risk reduction (CRR)¹ strategy is generally effective across setting, population, and program type on most of the stated outcomes, which include sexual activity, frequency of sex, number of partners, condom use, and STIs. Yet high measures of inconsistency of results across studies on key outcomes indicate that a sizable percentage of CRR programs did not work on these outcomes, especially for school-based programs, which are the focus of the national policy debate about sex education.

- a. The study suffers from a problem that is fundamental in meta-analysis by combining “apples and oranges.”²** The CRR category was created by collecting together everything that was *not* a study of “abstinence-only” sex education, resulting in a very heterogeneous category that exceeds the limits of good meta-analysis design. This category combines data from school classroom-based programs (a common delivery system for sex education that is fairly homogeneous with respect to setting, population, and pedagogy) with programs representing a wide variety of settings, populations, and pedagogies (e.g., STD clinics and youth shelters, youth in detention centers and parents in housing projects, individual service learning and multi-component youth development programs), illustrated in Table 1 (see Appendix). This problem undermines the validity of the entire study.
- b. When these studies of diverse programs were combined in the same analysis, the resulting statistical inconsistency in the data was not adequately resolved.** The Recommendation Statement does not report the important I^2 measure of heterogeneity for the meta-analysis results. (When a meta-analysis combines the effects from many studies into one measure to derive an average impact, the I^2 indicates how much inconsistency or variation there was across all of the effects that produced the average.) Yet heterogeneity for the majority of the significant CRR outcomes was so high that it would be considered “severe” by some experts,³ indicating that results reported at that level of aggregation are not interpretable. With such a level of internal inconsistency, the study recommendations should have indicated which of the diverse types of CRR programs in the study was effective and which was not, or reported their inability to do so, instead of concluding that there was sufficient

evidence of across-the-board effectiveness (“across a range of populations and settings”). This is misleading to policy makers seeking to choose evidence-based programs.

- c. **The small size of the reported CRR effects on key outcomes was not discussed in the Task Force Recommendations, and raises the question: What degree of behavioral improvement should a typical sex education program produce in order for it to be considered an “effective” program?** The relative risk estimates reported in the Recommendation Statement⁴ indicate that the CRR programs increased frequency of teen condom use by an average of 12% and reduced sexual activity by an average of 12%. They averaged a 16% improvement across the key outcomes of sexual activity, number of partners, condom use, pregnancy, and STIs (excluding conflated outcome measures, see below). This small magnitude of improvement, combined with the high inconsistency of results across studies, further diminishes the likelihood that any single CRR program will be effective at protecting adolescents. (It should be noted that the AE effects were in a similar range of magnitude.)
- d. **The Recommendation Statements do not report the lack of effectiveness for school-based CRR programs on key outcomes (condom use, pregnancy, sexually transmitted infections/STIs).**
 - 1. This is important since the school classroom is where most youth receive sex education and many of the programs that produced the reported general effects were not the school classroom type programs that most people think of as “sex education.”
 - 2. Nonetheless, the Recommendation Statement that “The evidence supports a conclusion that CRR interventions are applicable across a range of populations and settings...[including] school and community settings” gives policy-makers the impression that the study found CRR programs implemented in schools were effective at achieving these important outcomes, which was not the case.

2. Key measures of CRR program effectiveness are inadequate or confusing.

- a. **Instead of reporting on *consistent condom use* (CCU), the study reports on a less protective measure, *frequency of teen condom use*** (which combines studies that measured frequency of use, or use at first or last intercourse, with a few studies that measured consistent use). According to the CDC’s *Condom Fact Sheet*, “consistent and correct use of the male latex condom reduces the risk of sexually transmitted disease (STD) and human immunodeficiency virus (HIV) transmission. However, condom use cannot provide absolute protection against any STD. The most reliable ways to avoid transmission of STDs are to abstain from sexual activity, or to be in a long-term mutually monogamous relationship with an uninfected partner” (<http://www.cdc.gov/condomeffectiveness/brief.html>). STD (synonymous with STI) transmission can occur in one sexual contact and some studies have found that *non-consistent* use provided inadequate STD protection or resulted in higher rates of STDs.⁵ Rates of CCU by teens in the U.S. are low (28% of sexually active teen girls

and 47% of boys⁶) and this meta-analysis does not provide evidence that CRR programs have increased CCU and thereby increased actual protection.⁷

- b. The study uses a conflated outcome called *Use of Protection against Pregnancy and STIs*, which is a combined measure of condom use and oral contraceptive use.** It then reports a significant improvement on this outcome for CRR programs, implying that the programs were successful at affecting use of both condoms and oral contraceptives *and* that this was protective against pregnancy and STIs. **This is misleading because:**

1. It implies an effect on both condom use and oral contraceptive use, yet the study did not find significant effects for oral contraceptive use (when tested separately).⁴
2. Because oral contraceptives provide no protection at all against STIs, they should not be included in a measure of “protection against pregnancy *and* STIs”.
3. It is not known whether the condom/contraceptive use measured in this outcome *actually provided* “protection against pregnancy or STIs,” so to call it that is not accurate. (The study did not find that CRR programs generally reduced teen pregnancy.⁴)

For these reasons, *Use of Protection against Pregnancy and STIs*, is a misleading outcome; at the very least it should be renamed “Use of Condoms *or* Contraception;” to be most accurate the behaviors should not be combined in this measure.

- c. The study reports effects on another conflated outcome called *Unprotected Sex*, defined as having sex without using a condom.** A reduction in *Unprotected Sex* occurs when teens either abstain from sex *or* use a condom, thus this is a measure of whether sexual abstinence *or* condom use is occurring, without identifying the actual behavior. **This can be misleading and confusing to policy-makers:** combining both behaviors “pads” the numbers such that statistical significance can be obtained where it might not be obtained for both behaviors separately (in fact, this appears to have occurred for school-based CRR programs); it also gives CRR programs credit for effectiveness without identifying how that effect is being achieved. Finally, because studies of abstinence education have not typically measured *Unprotected Sex* as such, the CDC study **does not give AE credit for reducing “unprotected sex” by increasing abstinence**, although this is the effect of abstaining from sex.

- d. The study allowed minimum follow-up times for condom/contraceptive/STI outcomes (one to three months) that were too short to demonstrate a lasting program effect, but required longer follow-up times (six months minimum) for abstinence outcomes.** While not done with this intent, it had the effect of requiring most abstinence programs to meet a higher standard of effectiveness.

- 3. The meta-analysis evidence for the reported CRR effect on STIs is not of adequate quality to inform national policy about sex education.** The effect appears to be dependent upon only two CRR programs (DiClemente, 2004 and Jemmott, 2005).⁸ These programs were both clinic-based, i.e., not occurring in school classrooms with

school-based populations. In addition, evidence from several of the other studies in the analysis was of low quality.

- 4. The reduction in sexual activity showed by AE programs has been discounted because of a misplaced deference to randomized control trial (RCT) studies, some of which had important design problems.**

The meta-analysis found a significant reduction in sexual activity by the AE studies.⁴ However, the Task Force chose to designate it as “insufficient evidence” due to “inconsistent results across studies,” in the RCT (randomized control trials—the preferred research design) vs. non-RCT (or quasi-experimental, an accepted research design) results for sexual activity. **This decision is questionable for the following reasons:**

- a. Analyzing the results separately for the RCT and non-RCT studies assumes there were no other differences in research quality between these two sets of studies that might influence their outcomes.** Yet there were other differences, which resulted in some of the RCT studies having weaker research designs (see b. below).
- b. The RCT results were weighted heavily toward two studies by the same authors that have important design problems⁹** which likely caused them to inaccurately estimate a lack of abstinence effect. This raises questions about the deference given to their results by the Task Force.
- c. Of the 6 RCT studies in the AE analysis, 3 had sizable positive effects.**
- d. The measure of inconsistent results across studies was higher for CRR programs on sexual activity than for AE programs.**
- e. The quasi-experimental results are based on seven well-designed studies:** they met the criteria for inclusion in the CDC meta-analysis, the majority of them were peer-reviewed and published, and most used statistical methods to refine the comparison group match and control for pretest differences.

Thus, the meta-analysis provides credible evidence that AE programs have delayed sexual activity, evidence that is stronger than the evidence for CRR effects on STIs (see #3, above) or for the unreported evidence (i.e., lack of evidence) for school-based CRR programs on condom use (see #1d, above). This suggests an inconsistent standard has been applied. A case can be made that the difference in the RCT vs. non-RCT results for AE is due to problems in several RCT studies that prevented the accurate measure of positive results, not the opposite. This would mean that their influence on the AE results should be discounted rather than allowed to negate the body of research.

- 5. The Comparative Effectiveness Recommendation advocates CRR over AE based on assertions not supported by the meta-analysis evidence.**

- a. **The study sets up an inappropriate comparison between CRR and AE.** Given that school-based programs represented approximately 80% of the AE studies and only 40% of the CRR studies (see Table 1 in Appendix, including footnote), the appropriate comparison in the meta-analysis for determining relative effectiveness would have been between school-based CRR and AE programs. Yet the study conducted a comparative evaluation between AE, which was mostly school-based, and CRR as a whole, which was mostly community-based and consisted of very heterogeneous types of programs.
- b. The Task Force recommends CRR over AE because, it asserts, CRR has been shown to be “effective in both preventing sexual activity and reducing the risks of sexual activity [i.e., through the use of condoms/contraception].”⁴ However, **the meta-analysis methodology was not able to test empirically whether the CRR strategy has been effective at achieving both of these outcomes *within* individual CRR programs.** And if there is not evidence that the typical CRR program is effective at both outcomes then there is not evidence that it offers any real advantage over effective AE programs. While it should be acknowledged that this kind of evidence is hard to come by, it should also be recognized that lack of it constitutes *a lack of evidence*, and assertions that CRR achieves these dual outcomes should not be characterized as “evidence-based” where such evidence is lacking.
- c. **When the meta-analysis did a direct test of differences in effectiveness between the AE and CRR strategies, it failed to demonstrate a difference for any of the outcomes except one.**

In conclusion, we offer the above concerns and supporting rationale to explain our reasons for not supporting the Task Force Recommendations. Another version of this report, containing detailed statistical data from the meta-analysis as evidence for the above concerns, will be issued as soon as the CDC releases the meta-analysis data to the public. We would respectfully suggest that it is in the best interest of science and public policy for the CDC to release the meta-analysis data at the same time the Recommendation Statements are issued, so that the public can examine the full body of evidence upon which they are based—both supporting and otherwise. Lacking this, the above analysis has been presented to offer an alternative point of view on the interpretation of the study results.

About the authors:

Irene Ericksen, M.S., is a research analyst with *The Institute for Research & Evaluation* in Salt Lake City, a nonprofit research organization that has gained national recognition for its work evaluating sex education programs, particularly abstinence education. IRE has provided expert testimony to state legislative bodies, the U.S. Senate, the U.S. House of Representatives (April, 2008), and the White House (June, 2009). Ms. Ericksen is an author of peer-reviewed published research articles reporting the results of abstinence education evaluation studies.

Danielle Ruedt, MPH, is the Public Health Programs Coordinator for the Governor’s Office of Children and Families, State of Georgia.

Appendix

Table 1.

Type of Setting/Population^a	CRR studies	AE studies
A. School/Classroom setting, general school population	24 total (40%)	15 total (80%)
B. Community-based setting, self-selected population	36 total (60%)	4 total (20%)
Health or STD clinic setting & population	7	
Community center & population	6	3
Youth shelter & population	5	
Multi-component youth development program	4	
Parent training program (community-based)	4	1
Low-income housing project setting & population	3	
Youth in juvenile detention/jail	3	
Youth in residential drug treatment program	2	
Service learning by individuals at community sites	2	

- a. Note: This table was developed by the authors through examination of the list of studies included in the meta-analysis; it was not produced by the Community Guide research staff.

End Notes

1. The term “comprehensive,” when applied to sex education, is typically meant to indicate that interventions promote both sexual abstinence and sexual risk reduction (including the use of condoms and oral contraceptives). However, in this review of sex education studies, the comprehensive risk reduction (CRR) category includes studies of sex education programs that “solely promote sexual risk reduction,” i.e., that do not promote abstinence. This may be a source of confusion to policy-makers, who will miss this deviation from the accepted definition and assume that all of the programs in this category promoted teen abstinence to some degree.
2. See: Sharpe D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, 17, (8): 881–901, and Higgins JPT & Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21: 1539–1558.
3. Higgins and Thompson, 2002, give a magnitude for I^2 that indicates “severe heterogeneity” and conclude that I^2 values in this range should “induce considerable caution” regarding the results of a meta-analysis, p.1553. Thompson, 1994, and Moayyedi, 2004, argue that this level of heterogeneity in the effect renders the result meaningless at that level of aggregation; that the source of heterogeneity should be identified by examining subgroups and the results should only be interpreted at the subgroup level. (See Moayyedi P. (2004). Meta-analysis: Can we mix apples and oranges? *American Journal of Gastroenterology*, 2004: 2297–2301; Thompson SG. (1994). Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*, 309: 1351–1355.)
4. These facts are presented in the CDC Task Force Recommendation Statement for Comprehensive Risk Reduction, available at: <http://www.thecommunityguide.org/hiv/index.html>.
5. See Crosby RA, DiClemente RJ, Wingood GM, Lang D, Harrington KF. (2003). Value of consistent condom use: A study of sexually transmitted disease prevention among African American adolescent females. *American Journal of Public Health*; 93: 901–2.; Shlay JC, McCung MW, Patnaik JL et al. (2004). Comparison of sexually transmitted disease prevalence by reported level of condom use among patients attending an urban sexually transmitted disease clinic. *Sex Transm Dis*; 31(3):154–60.; Ahmed S, Lutalo T, Wawer M et al. (2001). HIV incidence and sexually transmitted disease prevalence associated with condom use: a population study in Rakai, Uganda. *AIDS*; 15(16):2171–9.; Grinsztejn B, Veloso V, Levi J, Velasque L, Luz P et al. (2009). Factors associated with increased prevalence of human papillomavirus infection in a cohort of HIV-infected Brazilian women. *International Journal of Infectious Diseases*, 13, 72–80.; Martin ET, Krantz E, Gottlieb SL, Magaret AS, Langenberg A, et al. (2009). A Pooled Analysis of the Effect of Condoms in Preventing HSV-2 Acquisition, ARCH INTERN MED/Vol 169 (13):1233–1240; Centers for Disease Control and Prevention. (2003). *Fact Sheet for Public Health Personnel—Male Latex Condoms and Sexually Transmitted Diseases*. National Center for HIV, STD, and TB Prevention. Atlanta, GA: U.S. Department of Health and Human Services (paragraph 4). Retrieved October 31, 2003 from www.cdc.gov/nchstp/od/latex.htm. According to the CDC, “inconsistent use, e.g., failure to use condoms with every act of intercourse, can lead to STD transmission because transmission can occur with a single act of intercourse” (CDC, 2003). A study in the journal *AIDS* (Ahmed et al., 2001) found, “Irregular condom use was not protective against HIV or STD and was associated with increased gonorrhea/chlamydia risk.” A Denver study (Shlay et al., 2004) reported that “when all condom users were compared with non-users ($N=126,220$), there was limited evidence of protection against specific STD.” But when consistent vs. inconsistent users were compared, the consistent users had significantly lower infection rates.
6. Franzetta, Kerry. et al. 2006. “Trends and Recent Estimates: Contraceptive Use Among U.S. Teens.” *Child Trends Research Brief* #2006-04.
7. Some will argue that these lesser measures may be indicators of behavior that might reduce the risk of STDs. But given that some studies have found non-consistent condom use was not protective (see note #5 above), we assert that to be recommended by such a major force in public health information dissemination

as the CDC, programs should not be called effective when they have not produced evidence of success at achieving this yet partially protective outcome of consistent condom use.

8. DiClemente RJ, Wingood GM, Harrington KF et al. Efficacy of an HIV prevention intervention for African American adolescent girls: a randomized controlled trial. *JAMA* 2004;292(2):171–9; Jemmott JB, III, Jemmott LS, Braverman PK, Fong GT. HIV/STD risk reduction interventions for African American and Latino adolescent girls at an adolescent medicine clinic: a randomized controlled trial. *Arch Pediatr Adolesc Med* 2005;159(5):440–9.
9. Design Problems of RCT Abstinence Studies
 - a. The first of these studies was a long-term evaluation of 4 different abstinence programs, in which no positive effects were found. However, when reviewed critically, important design weaknesses become evident. Each of the 4 sub-studies randomized the assignment to treatment and control groups *at the individual level within schools*. The programs were then implemented within these schools for substantial periods of time, ranging from 1 to 4 years. This design sets up a classic scenario in which contamination or cross-fertilization of the treatment effect between the treatment and control groups can occur within the school (the “lunch-room-to-locker-room effect”). The result is a muting of the treatment effect over time, in that the two groups influence each others’ attitudes and behaviors in the close quarters of the school environment. This is especially likely when the treatment occurs over an extended time period. This, combined with the unusually young age of the program participants (8 to 13 years old) and the unusually long follow-up period used in the study (4 to 6 years after baseline) constitutes good reason to question this study as a valid test of abstinence education. (See Trenholm C, Devaney B, Fortson K, Quay L, Wheeler J, Clark M. (2007). *Impacts of Four Title V, Section 510 Abstinence Education Programs*. Princeton, NJ: Mathematica Policy Research, Inc. April 2007.)
 - b. The problem with the second study, Clark, Trenholm, 2007, has to do with its inclusion in the CDC meta-analysis at all. The study was included ostensibly because it was considered an evaluation of an abstinence curriculum. However, the purpose of this study was to detect whether the Heritage Keepers “Life Skills” curriculum had any incremental impact on teen sexual behavior over and above the Heritage Keepers Abstinence Curriculum. The Life Skills curriculum, which contained minimal abstinence content, was a voluntary after-school program intended as a supplement to the main Abstinence Curriculum, which was a mandatory class taught during the school day. In this study, the treatment group received the Abstinence Curriculum plus the Life Skills Component, and the control group received only the Abstinence Curriculum. Thus, this study did not test the effect of abstinence education versus the absence of abstinence education, but rather it tested the effect of a voluntary life skills program (i.e., a light abstinence dose) over and above the effect of a mandatory abstinence program (i.e., a heavy dose of the abstinence message). In short, it was not an evaluation of the effectiveness of an abstinence education program, and as such, should not have been included in the group of abstinence studies in the meta-analysis. It should be noted that the Heritage Keepers Abstinence Curriculum was evaluated against a true counterfactual (the absence of abstinence education) and demonstrated a significant and sizable reduction in teen sexual initiation (OR=.54, $p<.001$) one year after the program. (See Clark MA, Trenholm, C, Devaney, B, Wheeler, J, Quay, L. (2007). *Impacts of the Heritage Keepers Life Skills Education Component. Final Report*. Princeton, NJ: Mathematica Policy Research, Inc. and Weed SE, Ericksen IH, Birch PJ. (2005). *An evaluation of the Heritage Keepers Abstinence Education program*. In Golden A (Ed.) *Evaluating Abstinence Education Programs: Improving Implementation and Assessing Impact*. Washington DC: Office of Population Affairs and the Administration for Children and Families, Department of Health & Human Services 2005:88–103.)
 - c. The above concerns constitute good reason to question these studies as a valid test of abstinence education and, given their heavy weight in the analysis, are good reason to question the RCT versus non-RCT results for the abstinence effects in the meta-analysis.